
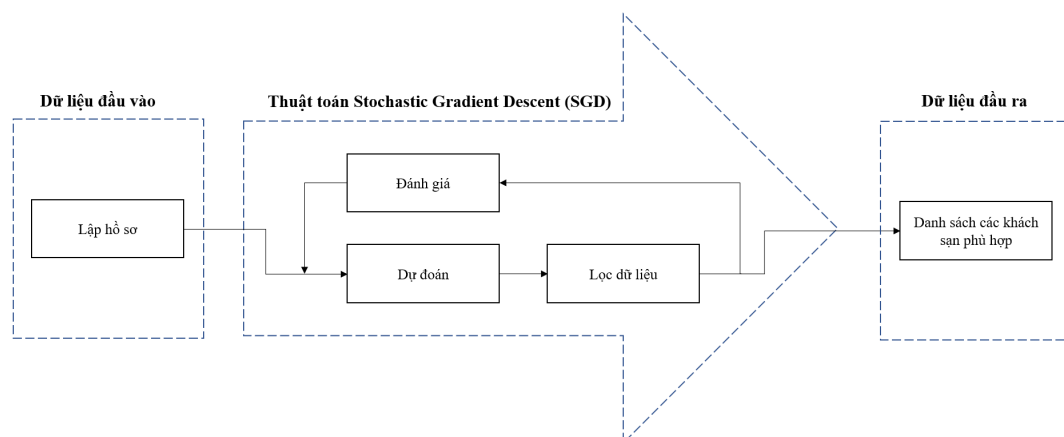


MẪU BÁO CÁO CỦA MỖI HV

Họ và tên (IN HOA)	NGUYỄN THANH PHONG MSHV: CH2002014
Ảnh	
Số buổi vắng	1
Bonus	16
Tên đề tài (VN)	LẬP HỒ SƠ KHÁCH HÀNG VÀ GỢI Ý KHÁCH SẠN
Tên đề tài (EN)	ON-LINE GUEST PROFILING AND HOTEL RECOMMENDATION
Giới thiệu	<ul style="list-style-type: none">• <i>Bài toán/vấn đề mà đề tài muốn giải quyết</i><ul style="list-style-type: none">- Gợi ý khách sạn đến khách du lịch dựa trên xếp hạng, đánh giá do nguồn cộng đồng cung cấp và thông tin khách sạn trực tuyến.• <i>Lí do chọn đề tài, khả năng ứng dụng thực tế, tính thời sự</i><ul style="list-style-type: none">- Công nghệ thông tin và truyền thông (ICT) hiện nay đã cách mạng hóa ngành du lịch, cụ thể công nghệ di động có ảnh hưởng trực tiếp đến trải nghiệm du lịch của khách du lịch (KDL) do tính phổ biến và tiện lợi của nó về việc truy cập internet, dịch vụ và ứng dụng.

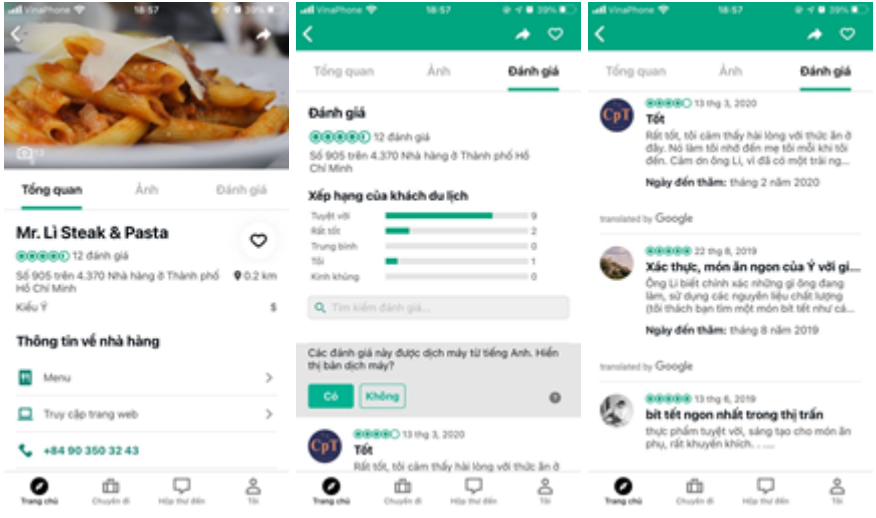
- Sự tương tác thường xuyên của KDL và ICT tạo ra khối lượng lớn dữ liệu có nguồn gốc từ cộng đồng. Thông tin gồm (Xếp hạng, lượt thích, lượt chia sẻ và đánh giá về địa điểm du lịch) -> Tạo ra hiệu ứng gợn sóng khi quyết định của KDL hiện tại và tương lai ảnh hưởng rất nhiều bởi ý kiến từ những người đi trước.
- KDL chia sẻ các thông tin như Nhận xét, đánh giá, xếp hạng, video, ảnh, bài đăng hoặc lượt thích) lên các nền tảng du lịch như TripAdvisor, Expedia hoặc Booking.com (ảnh minh họa nguồn trích dẫn từ TripAdvisor)
- Dữ liệu này rất quan trọng với KDL và doanh nghiệp (DN). KDL dựa trên các ý kiến phản hồi của người đi trước để từ đó đưa ra quyết định lựa chọn dịch vụ. DN thì khai thác dữ liệu phản hồi của KDL để xác định vấn đề tồn tại và xu hướng, bằng cách áp dụng thuật toán để dự đoán hành vi, sở thích của KDL.
- Bài báo này, tác giả đề xuất một công cụ đề xuất dựa trên xếp hạng, đánh giá do nguồn cộng đồng cung cấp và thông tin chính thức từ khách sạn (KS).

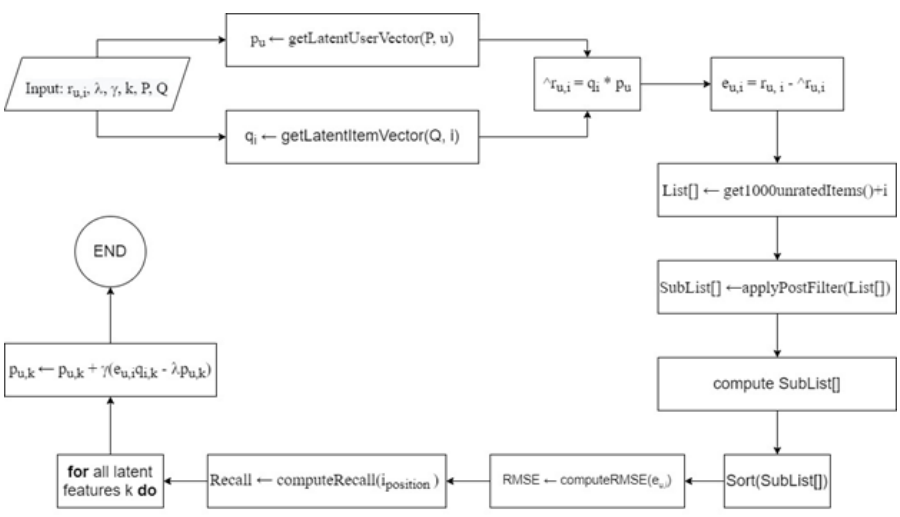
● *Mô tả input và output, nên có hình minh họa*



- Dữ liệu đầu vào:

Tập dữ liệu HotelExpedia, bao gồm 6.276 khách sạn và 1.090 người dùng, 214.342 đánh giá từ 11 địa điểm khác nhau. Mỗi người dùng đã

	<p>xếp hạng ít nhất 20 khách sạn và mỗi khách sạn có tối thiểu 10 xếp hạng.</p>  <p>The screenshots show a mobile application interface for a restaurant named 'Mr. Li Steak & Pasta'. The first screenshot shows the restaurant's profile with a 4.5-star rating from 12 reviews, location at 905 Trien 4.370, and a distance of 0.2 km. The second screenshot shows a detailed review from a user named 'Cpt' dated 13th Aug 2020, with a 5-star rating and a positive comment about the food. The third screenshot shows another review from a user named 'Cpt' dated 13th Aug 2020, with a 5-star rating and a positive comment about the food.</p> <ul style="list-style-type: none"> - Dữ liệu đầu ra: <p>Danh sách các khách sạn phù hợp với tiêu chí của khách du lịch (KDL).</p>
Mục tiêu	<ul style="list-style-type: none"> - Bài báo khai thác thông tin có nguồn gốc từ cộng đồng về ngành khách sạn bằng cách sử dụng dữ liệu của Expedia, nhiều xếp hạng tiêu chí và đánh giá văn bản cũng như thông tin chính thức của khách sạn - Bài báo là một công trình nghiên cứu, cung cấp cách tiếp cận hồ sơ mới về khách và khách sạn dựa trên thông tin có sẵn từ nguồn cộng đồng – xếp hạng đa tiêu chí, đánh giá bằng văn bản và thông tin khách sạn - Bài báo cho thấy thuật toán phân tích nhân tử ma trận Stochastic Gradient Descent (SGD) là một kỹ thuật đầy hứa hẹn, khi được kết hợp với các cấu hình đã được tinh chỉnh và các bộ lọc sau đề xuất thích hợp, rất phù hợp cho việc cung cấp trực tuyến khuyến nghị của khách sạn.
Nội dung và	<ul style="list-style-type: none"> • <i>Viết chi tiết các nội dung và phương pháp để đạt mục tiêu</i> <ul style="list-style-type: none"> - Bước 1: Khởi tạo ma trận xếp hạng với các xếp hạng khách hàng.

<p>phương pháp thực hiện</p>	<ul style="list-style-type: none"> - Bước 2: Xây dựng ma trận tiềm ẩn cho khách hàng và khách sạn, phân phối ngẫu nhiên một thành phần phạm vi nhỏ từ -0,02 đến 0,02 để đảm bảo các hệ số là khác nhau. - Bước 3: Tạo ra các dự đoán bằng cách sử dụng ma trận tiềm ẩn. Sau đó, tác giả chọn ngẫu nhiên 1000 khách sạn không được xếp hạng cộng với khách sạn mới được xếp hạng, tức là 1001 khách sạn. Thuật toán dự đoán xếp hạng của 1001 khách sạn đã chọn và sắp xếp chúng theo thứ tự giảm dần.
<p>Kết quả dự kiến</p>	<ul style="list-style-type: none"> • <i>Thuật toán SGD (Dựa trên dữ liệu khách sạn)</i> <ul style="list-style-type: none"> - Mô hình thuật toán  <p style="text-align: center;">(Độ phức tạp của thuật toán là $N \log N$.)</p> <ul style="list-style-type: none"> - Mã giả: <ol style="list-style-type: none"> 1: Input: $\neg r_{u,i}, k, P, Q$ 2: Output: $\hat{r}_{u,i}$ 3: $p_u \leftarrow \text{getLatentUserVector}(P, u)$ 4: $q_i \leftarrow \text{getLatentItemVector}(Q, i)$ 5: $\hat{r}_{u,i} = q_i * p_u$

```

6:  $eu, i = ru, i - \hat{ru}, i$ 
7:  $List[] \leftarrow get1000unratedItems() + i$ 
8:  $SubList[] \leftarrow applyPostFilter(List[])$ 
9: for  $i \leftarrow SubList []$  do
10:  $q \leftarrow getLatentItemVector(Q, i)$ 
11:  $\hat{ru}, i = q_i * p_u$ 
12:  $Sort(SubList[])$ 
13:  $RMSE \leftarrow computeRMSE(eu, i)$ 
14:  $Recall \leftarrow computeRecall(iposition)$ 
15: for all latent features  $k$  do
16:  $p_{u,k} \leftarrow p_{u,k} + (eu, i q_{i,k} - p_{u,k})$ 

```

- So sánh giữa các phương pháp

Table 1

Comparison of hotel recommendation research approaches.

Approach	Evaluation	Profiling	Prediction	Post-Filtering
Song et al. (2016)	TripAdvisor	Rat	ED	
Farokhi et al. (2016)		Rat	<i>k</i> -means	
Dong and Smyth (2016)		Rev	Similarity	
Shrote and Deorankar (2016)		Rev	SA	
Ebadi and Krzyzak (2016)	TripAdvisor	Rat & Rev	SVD & TP	Loc, StV & Vfm
Sharma et al. (2015)	TripAdvisor	Rat & Rev	NLP	
Hu et al. (2016)	TripAdvisor	Rev	TF-IDF	
Hariri et al. (2011)	TripAdvisor	Rev	CS + LDA	
Current proposal	Expedia	Rat & Rev	SGD	

- Bộ dữ liệu, etc
 - Tập dữ liệu HotelExpedia

Tài liệu tham khảo	<p>[1] Amatriain, X., 2013. Mining large streams of user data for personalized recommendations. ACM SIGKDD Expl. Newslett. 14 (2), 37–48.</p> <p>[2] Aramendia-Muneta, M.E., Ollo-Lopez, A., 2013. Ict impact on tourism industry. Int. J.Manage. Cases 15 (2), 87–98.</p> <p>[3] Book, L.A., Tanford, S., Chen, Y.S., 2016. Understanding the impact of negative and positive traveler reviews: Social influence and price anchoring effects. J. Travel Res. 55 (8), 993–1007.</p> <p>[4] Borràs, J., Moreno, A., Valls, A., 2014. Intelligent tourism recommender systems: a survey. Expert Syst. Appl. 41 (16), 7370–7389.</p> <p>[5] Cremonesi, P., Koren, Y., Turrin, R., 2010. Performance of recommender algorithms on top-n recommendation tasks. In: Proceedings of the fourth ACM conference on Recommender systems ACM, pp. 39–46.</p> <p>[6] Dong, R., Smyth, B., 2016. From more-like-this to better-than-this: Hotel recommendations from user generated reviews. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization ACM, pp. 309–310.</p>
-----------------------------------	--